Chapter 19
Discussion of three resampling papers

Peter J. Bickel

Abstract
Discussion of:
Putter H., and van Zwet W.R. (1996). Resampling: Consistency of substitution estimators, *Annals of Statistics* **24** 2297–2318.
Putter H., and van Zwet W.R. (1997). On a set of the first category. *Festschrift for Lucien Le Cam, Springer Verlag*, 315–324.
Bickel P., Götze F. and van Zwet W.R. (1997). Resampling fewer than n observations: Gains, Losses and Remedies for Losses, *Statistica Sinica* **1** 1-31.

It is a pleasure to return to these three papers of van Zwet's on the bootstrap, two coauthored with Hein Putter and the other with Friedrich Götze and myself.

They marked van Zwet's attempt to understand the behaviour of bootstrap estimates of parameters, when observations, $X_1,\ldots,X_n$ were i.i.d. for $P \in \mathscr{P}$. This was done for clarity of conception only. It was evident that generalizations to weakly dependent data should hold.

In the first two papers Putter and van Zwet considered a sequence $\tau_N(P)$ of parameters which were themselves probability distributions, and endowed with a suitable metric (e.g. Prohorov-Levy). $\mathscr{P}$ was endowed with a metric ρ. If P_N is a ρ consistent estimate of P and

$$d\big(\tau_N(\hat{P}_N), \tau_N(P)\big) \xrightarrow{P} 0 , \tag{19.1}$$

the bootstrap $\tau_N(\hat{P}_N)$ was defined as successfully estimating $\tau_N(P)$.

Their main emphasis in the first paper was to show the general feasibility of constructions satisfying (refequation1) for continuous τ_N except on sets of the first category in $\mathscr{P}$. This led then, on the one hand, to the parametric bootstrap and, on the other, to the view that bootstraps and $\hat{P}_N$ had to be tailored to the specific problem, for interesting τ_N. Essentially this is the case unless ρ is the Hellinger

Peter J. Bickel
Department of Statistics, 367 Evans Hall. Berkeley, CA 94710-3860
e-mail: bickel@stat.berkeley.edu

S. van de Geer and M. Wegkamp (eds.), *Selected Works of Willem van Zwet*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1314-1_19, © Springer Science+Business Media, LLC 2012

metric in which case, ρ consistency of any $\hat{P}_N$ typically fails unless $\mathscr{P}$ is regular parametric – but see Golubev et al [1].

The second paper gave interesting examples of situations where such $\hat{P}_N$ could be constructed but, unless a great deal of attention to the structure of $\mathscr{P}$ was given, would have the set of the first category mentioned be all of $\mathscr{P}$!

The third paper jointly with Götze and myself took a different track, concentrating on sequences, $\tau_N(P,\hat{P}_N)$ such as those that usually arise in setting confidence bounds and other statistical questions, e.g. $\tau_N(P_N,P) = N^{-\frac{1}{2}}\left(\int x d\hat{P}_N - \int x d\hat{P}\right)$ for $\mathscr{P} = \left\{P : \int x^2 dP < \infty\right\}$.

Here, and throughout the sequel, $\hat{P}_N$ is the empirical distribution. As in the above example, the τ_N are assumed to have a basic property quite different from the τ_N considered by Putter and van Zwet:

$$\text{(I)} \qquad \tau_N(\hat{P}_N,P) \text{ converges weakly to a limiting probability distribution } \tau_P\ .$$

In the example above $\tau_P = \mathscr{N}\left(0, \mathrm{Var}_P(X_1)\right)$. Our focus was on estimating the probability distribution of $\tau_N(\hat{P}_N,P)$.

Condition I trivially implies that, if $m_N \to \infty$, $\tau_{m_N}(\hat{P}_{m_N},P)$ converge weakly to τ_P as well. The condition suggests that we use implicit scaling as in the example and estimate the distribution of $\tau_N(\hat{P}_N,P)$ by that of $\tau_{m_N}(\hat{P}^*_{m_N},\hat{P}_N)$ where $m_N \to \infty$ slowly, $\{\hat{P}^*_m\}$ depends on $\hat{P}_N$ and converges to P in an appropriate metric ρ.

Independently, Politis and Romano (1994) and Götze (1993) considered the basic but statistically less interesting case that $\tau_N(\hat{P}_N,P) \equiv \tau_N(P)$, so that τ_p is degenerate. They showed that if $\hat{P}^*_m$ is the empirical distribution of a sample drawn *without replacement* from $X_1,\ldots,X_N$, $m_N \to \infty$ and $\frac{m_N}{N} \to 0$, then the conditional distributions of $\tau(\hat{P}^*_m,\hat{P}_N)$ given the data converge weakly, with probability 1, to τ_P without any further conditions. Our (1997) paper goes on to investigate when $\tau_{m_N}(\hat{P}^*_{m_N},\hat{P}_N)$ converges to τ_P generally, both when $\hat{P}^*_{m_N}$ corresponds to sampling without replacement and with replacement. A number of other issues are also studied. Not surprisingly when $\tau_N(\hat{P}^*_N,\hat{P}_N)$ converges weakly to τ_P, it typically does so faster than $\tau_{m_N}(P^*_{m_N},\hat{P}_N)$ with $\frac{m_N}{N} \to 0$. We discuss ways of removing this disability and propose a crude rule for data determined selection of m_N.

This approach and general applications to situations where the Efron bootstrap fails are analyzed further in Götze and Rakauskas (2001), Bickel and Sakov (2002a) and Bickel and Sakov (2002b).

The papers with Putter exhibit van Zwet's typical approach to research: A general question is sharply posed followed by a definitive, technically subtle answer, in this case, I think, not as satisfactory as van Zwet originally hoped. The 1997 paper though considerably less elegant and definitive than the work with Putter would seem to have the general applicability that van Zwet initially hoped for – but that's obviously a biased opinion.

References

1. Golubev G.K., Levit B., Tsybakov, A.B. (1996). Asymptotically efficient estimation of analytic functions in Gaussian noise, *Bernoulli, 2*, 167–181.
2. Götze F. and Rakauskas, A. (2001). Adaptive choice of bootstrap sample sizes, *State of the Art in Probability and Statistics*, 286–309.
3. Bickel P. and Sakov A. (2002). Extrapolation and the bootstrap, *Sankhya, Ser. A, 3*, 640–652.
4. Bickel, P. and Sakov, A. (2002). On the choice of m in the m out of n bootstrap and confidence bounds for extrema, *Statist. Sinica, 18*, 967–985.
5. Politis, D. and Romano, J. (1994). Large sample confidence regions based on subsamples under minimal assumptions, *Ann. Statist. 22*, 2031–2050.
6. Götze F. (1993). Asymptotic approximations and the bootstrap, *JMS Bulletin*, 305.